Landbird Monitoring Protocol for Klamath Network Parks

Standard Operating Procedure (SOP) #13:
Data Validation and Verification

Version 1.0

Revision History Log:

Previous Version	Revision Date	Author	Changes Made	Reason for Change	New Version

This SOP explains the procedures for verifying and validating data and includes information on handling field forms.

Field Forms

The field data forms are part of the permanent record; they should be handled in a way that preserves their future interpretability and information content. If changes to data on the forms need to be made after the day of data collection, the original values should not be erased or otherwise rendered illegible. Instead, changes should be made as follows:

1. Edits and revisions should be made in red pencil to make it easier for subsequent viewers to be able to retrace the edit history.
2. A horizontal line should be drawn through the original value and the new value written adjacent to the original value with the date and initials of the person making the change.
3. All corrections should be accompanied by a written explanation in the appropriate notes section on the field form. These notes should also be dated and initialed.
4. These procedures should be followed throughout data entry and data validation, until the final annual dataset is created. After that point, edits are made only in digital edit logs and not on the hardcopies of the data.

Data Validation and Verification

Technician and/or Intern

The data undergo two rounds of review by the technician and/or intern. The goal of data proofing and editing by the technician and/or intern is to ensure all data have been recorded and to make certain each record in the database matches its corresponding record on the datasheet. The first data review occurs in the field at the end of each survey, when the observer proofs and edits the data. This proofing and editing consists of looking for typos, obvious mistakes, and missing data. If additional data have been recorded in a field notebook, they are entered on the field forms at the end of the day. For example, during mist netting, birds are occasionally released or escape at

SOP #13: Data Validation and Verification (continued).

the net with capture information recorded in field notebooks. The minimal data gathered (i.e., species, sex, net number) are then recorded on datasheets at the end of the banding session. After each page of data is proofed and edited, the observer initials the bottom of the datasheet to indicate completion.

The second round of proofing is completed by the technician and/or intern when the data are entered into the Access databases (SOP #12: Data Entry). The Access data entry interface has domain values built in on all fields to increase the accuracy of data entry. For example, if the observer enters a bird code that is not in the bird list, an error message is displayed regarding the problem. The person entering the data is allowed to correct minor errors while entering data (e.g., misspelling a species code or improperly ageing or sexing a bird when the data unambiguously supports a different determination). When minor errors or typos on the datasheets are encountered during data entry, they are corrected with red pencil and noted in the "Error Notes" field. In addition, all corrections are logged in a data entry log that corresponds with each data file (SOP #12: Data Entry). After completion of data entry, all unresolved errors are forwarded to the Project Lead for further scrutiny.

Project Lead
Two additional rounds of review are completed by the Project Lead. The first occurs when datasheets and Access databases are forwarded by the technician and/or intern to the Project Lead. The Project Lead reviews the unresolved errors identified during the data entry process. When possible, these errors are corrected with red pencil and noted on the field forms. In addition, all corrections are made in the Access database, noted in the "Error Notes" field, and logged in the text-based log file that corresponds with each data file (SOP #12: Data Entry).

The second round of review, completed by the Project Lead, involves data validation. Data validation is the process of checking data for completeness, structural integrity, and logical consistency. The Project Lead compiles the data from all observers into a single dataset for each survey type (e.g., point count, mist net). Each of these files has a root name that includes two digits indicating the name of the survey method, two digits indicating the project, and two digits indicating the year the data were collected (e.g., VRKB06, BDKB06). The Project Lead then completes his/her second round of proofing and editing. During this process, the Access databases are backed up each time data validation and related editing occurs. The backup files are named with the root file name followed by the date (e.g., VRKB06 20060131).

A series of validation programs, written in SAS, R, and Visual Basic, produce outputs that allow the Project Lead to identify errors in the data. The programs produce a series of error reports, which are then checked by the Project Lead. All error reports are stored electronically with the databases and hardcopies printed and filed. For point count, area search, and vegetation data, error reports include:
1. Number of stations and visits per route – The Project Lead compares this output to the actual number of stations and visits that were to be completed. If the numbers do not match, he or she compares the database to the field forms to determine where the error occurred. This process identifies errors in route names, station numbers, or survey dates.

SOP #13: Data Validation and Verification (continued).

2. Routes and stations missing UTM coordinates – This output contains gaps where the UTM coordinates are missing. The Project Lead compares the dataset and the UTM database to identify errors in route names or missing location information.
3. Species codes and common name with number of detections – This output allows the Project Lead to identify incorrect codes that do not output with a common name and suspect codes of uncommon species. These questionable records are cross-referenced with field forms to determine appropriate edits. If necessary, observers are contacted to provide verification.

For banding data error reports, include:

1. Incorrect age and sex determinations that are both inappropriate for a particular time of year and/or for individual species and measurements that falls outside of known ranges for each species.
2. Duplicate records.
3. Incorrect band numbers.
4. Recaptures marked as newly banded or unbanded and recaptures of birds at different stations.
5. Data that do not correspond to previous years' data (e.g., duplicate band numbers, bands incorrectly recorded as new captures when they are actually recaptures, and band numbers recorded at different stations).

After all edits have been identified by validation programs, minor errors (e.g., route names, station numbers, improper species codes, minor ageing and sexing errors) are corrected, noted in the "Error Notes" field, and logged into the database log file for the database. This process results in a database that matches the datasheets, error reports, and log of database edits.

Once the validation and verification steps have been completed, the databases are uploaded into the KLMN Master Bird databases where they receive another round of data validation. At this stage, the datasets are archived and made ready for distribution and analyses.

Editing Archived Datasets
As datasets are analyzed, additional edits may be identified. Any further editing must follow the editing archived datasets procedure detailed below:

1. Ensure that the dataset is backed up correctly, as detailed above in the data verification section.
2. Maintain a digital data entry log with meticulous notes tracking all edits and the reason for needed changes.
3. Incorporate edits into the Access dataset with notes in the "Error Notes" field and output a revised dataset.
4. Resulting datasets should be checked for accuracy of edits and retention of all records.

SOP #13: Data Validation and Verification (continued).

5. The old dataset can be easily recovered if editing was incorrect by returning to the dataset backed up in step 1.
6. Transfer the revised dataset, along with the digital data management log, to the NPS Data Manager, to be incorporated into the KLMN Landbird database (SOP #17: Data Transfer, Storage, and Archiving).